AD_____

Award Number: DAMD17-01-1-0800

TITLE: Identification of Alcoholism Susceptibility Genes

PRINCIPAL INVESTIGATOR: Kirk C. Wilhelmsen, M.D., Ph.D.

CONTRACTING ORGANIZATION: Ernest Gallo Clinic and Research Center
Emeryville, California

REPORT DATE: October 2002

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2002	Annual (15 Sep 01 - 15 Sep 02)

4. TITLE AND SUBTITLE
Identification of Alcoholism Susceptibility Genes

5. FUNDING NUMBERS
DAMD17-01-1-0800

6. AUTHOR(S)
Kirk C. Wilhelmsen, M.D., Ph.D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Ernest Gallo Clinic and Research Center
Emeryville, California

E-Mail: kcw1@itsa.ucsf.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Genetic epidemiological analysis provides convincing data that a significant portion of the liability for developing alcoholism is inherited. The identity and mechanism by which genes contribute to inherited susceptibility to alcoholism are unknown. If genes that affect susceptibility to alcoholism can be identified, they are logical targets for the development of pharmacological agents to modify susceptibility and treat alcoholism. Primarily, the meiotic gene mapping approach has been used in an effort to identify genes that affect susceptibility to alcoholism. This approach may be insensitive to genes that have common alleles with modest effects on susceptibility. An alternative approach, allelic association analysis, may be far more powerful for detecting these genes, but requires the analysis of individual candidate genes. This proposal examines a large number of genes implicated in the biology of alcoholism to see whether common alleles of these genes affect susceptibility. By Nov 30, 389 polymorphisms, from 23 candidate genes, in over 1200 subjects from the UCSF family study will have been genotyped. The advantage of performing this analysis on a large number of genes is that the data can be used to avoid population stratification and may detect allelic effects on susceptibility that would not otherwise be observed.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Alcoholism, Allelic Association, Human Genetics, Susceptibility, Single Nucleotide Polymorphism, Automation	10
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Genetic epidemiological analysis provides convincing data that a significant portion of the liability for developing alcoholism is inherited. The identity and the mechanism by which genes contribute to inherited susceptibility to alcoholism are unknown. If genes that affect susceptibility to alcoholism can be identified, they are logical targets for the development of pharmacological agents to modify susceptibility and treat alcoholism. The bulk of the effort that has been made to identify genes that affect susceptibility to alcoholism has used the meiotic gene mapping approach. This approach, however, may be insensitive to genes that have common alleles that have modest effects on susceptibility to alcoholism. Allelic association analysis is an alternative approach that may be far more powerful for detecting these genes, but requires the analysis of individual candidate genes. This is a proposal to examine a large number of genes implicated in the biology of alcoholism to see whether common alleles of these genes affect susceptibility. The infrastructure created in this proposal will be scaleable such that ultimately a substantial fraction of the genes in the human genome could be scanned. The advantage of performing this analysis on a large number of genes is that the genotypic data can be used to detect and avoid population stratification and may allow for the detection of allelic effects on susceptibility that would not otherwise be detected.

Progress Report Body

The principal technical objective of this proposal is to develop the infrastructure that can be scaled to perform a genome-wide allelic association analysis. The short-term goal is to efficiently screen a large number of candidate genes for allelic association with alcoholism. As part of this process, algorithms have been implemented and improved so that assay optimization and genotype determination can be accelerated. The proposed study is outlined here.

1) Develop a database application to organize this project. This database incorporates the following features (Note that many of these features are an extension of our previously developed database used for our high throughput microsatellite genotyping database application):
    a) Track the data supporting candidate gene selection including a workspace to record links to primary literature and conclusions reached.
    b) Tools for automatically periodically querying to find updated sequence and new SNP data for candidate genes.
    c) Tables for storing information concerning selected SNPs (e.g., Local sequence, coding vs. non-coding etc).
    d) Tools for batch designing PCR primers and probes.
    e) Tables to record sequence information for selected probes and oligonucleotide-specific information (e.g. storage location, concentration, and amount remaining etc).
    f) Tools for generation of automated pipetting protocols for PCR and assay optimization.
    g) Tables for storage of results of optimization experiments.
    h) Tools for generation of automated pipetting protocols for genotyping.
    i) Tables for storage of fluorimetry data.

j) Tools for analyzing fluorimetry data including an assessment of reliability for repeated samples.
k) Tables for storage of genotypes.
l) Links to sample and clinical phenotype and pedigree databases.
m) Tools for formatting data for allelic association analysis.
n) Tables for storing the results of association analysis.

2) Screen at least 100 candidate genes for allelic association using an average of 6 SNPs per gene in a population of 1200 subjects over the next four years. The genes that will be screened in the first year are listed in Table 4.
3) Use the average genotype sharing for markers for an individual relative to a population distributed throughout the genome to identify population stratification and outliers that can interfere with the detection of allelic association.
4) Determine if phenotypic elements of the diagnosis of alcoholism are responsible for observed allelic associations.

Since this proposal was submitted there have been a number of developments in the field of human genetics that confirm the design of this project. The most significant change has been the recognition of the structure of linkage disequilibrium in the genome. It has been demonstrated by many groups that regions of a few kb where recombination occurs separate regions of between 10-100 kb that are resistant to recombination. There are limited numbers of common haplotypes for the recombination resistant regions in entire population of the world. This observation has profound implications for this proposal and is discussed below.

Since we have the greatest power to detect allelic association for traits that are common in the population we have refocused our efforts to develop a scheme that allows us to conclusively determine whether common alleles of the genes screened are in linkage disequilibrium with the trait. In the proposal we planned to screen a small number of single nucleotide polymorphisms (SNPs) for each gene as a survey. These SNPs were selected based on the probability that they could affect the function of the gene and whether the SNPs were likely to be informative. The theory was that association between the gene and the trait would be detected even if the SNPs directly assayed were not responsible for the trait, because they could be in disequilibrium with the causal SNP(s). Current research on linkage disequilibrium supports the hypothesis that we should be able to genotype a smaller number of SNPs for each gene to accomplish a more systematic screen, provided that the most useful SNPs are selected. To take full advantage of linkage disequilibrium, we have added a single step that will allow us to determine which SNPs

are the most informative for haplotype determination and the optimal number of SNPs to genotype for a given gene. Ultimately the addition of this step will effectively streamline the process, as much fewer SNPs will need to be assayed per gene in the population under analysis to obtain the same amount of allelic information. The basic strategy is to resequence the regions conserved between humans and mice for a collection of 96 individuals. The 96 individuals will be from 32 families with a child and both parents. The sequence of these individuals will give us 128 haplotypes. We will then select the most informative SNPs to genotype so that all of the common haplotypes within a gene can be identified.

To facilitate this haplotype determination process we have constructed informatics tools to annotate the available genome sequence and organize all of the data generation steps and analysis that will be done. We have successfully designed and incorporated tools to identify the conserved sequences between species, collect information on all SNPs previously identified in all public and the Celera database (including allele frequency and location within the gene), prioritize the SNPs and regions for resequencing. We are currently incorporating primer design tools and tools for developing automated protocols for sample handling for both sequencing and SNP assay assembly.

We currently are at the stage in which the central laboratory database is an integral part of all work in the lab. All blood specimens and DNA samples are bar coded. All racks, storage and assay plates are bar coded. All major processes in the laboratory are done with the assistance of a database interface that facilitates the recording of activity. For example, we have developed robotic protocols to measure the concentration of DNA in samples. The application allows the user to scan sample tubes containing stock DNAs and place them in a rack on a pipetting robot. The robot fills a 96 well optical grade plate with dilutions suitable for direct DNA concentration determination. The computer makes a template for the spectrophotometer. The output file from the spectrophotometer is read by the database so that the results are stored in the database. Using this DNA concentration data, automated pipetting protocols are also written by the computer to make dilutions and set up assays for these DNA samples.

The stages in screening a candidate gene for association with alcoholism are: 1) selection of genes, 2) the selection of sequences to be screened for polymorphisms, 3) development of amplimers, 4) production of sequence, 5) review and analysis of sequence data for quality and polymorphism identification, 6) determination of haplotypes, 6) and identification of the SNPs to be genotyped in the extended population, 7) development of assays, 8) production of genotypes for the extended population, 9) haplotype based association analysis, and 10) systematic analysis of all the polymorphisms in candidate genes with positive results. We have made substantial progress on developing the tools needed for steps 1,2,4,5,6 and 9. The most ambitious advancement has been for step 5 in which we have extended the analysis tools available for analyzing sequence, to include automated genotype determination and verification of polymorphisms. All of these analysis tools are used in conjunction with the master database.

During the development of the tools described above we have proceeded with candidate gene association analysis. The SNPs under analysis come from 23 candidate genes (Table 1) that have been implicated in the biology of alcoholism from studies conducted in model systems. By the end of November, 2002 we will have completed genotype determination of 389 SNPs in over 1200 subjects from the UCSF family alcohol study. Once the haplotyping step described above is implemented, fewer SNPs per gene will need to be analyzed per gene. However, the larger number of SNPs assayed here per gene will provide us with an enhanced SNP dataset with which we can compare allelic frequency with our study population with that found in the population analyzed as part of other collections, such as the Celera SNP database. This dataset will allow us to determine whether any of these genes have common variations that affect susceptibility to alcoholism.

In addition, linkage analysis studies in my laboratory have confirmed related studies done by others and have suggested that several chromosome regions contain loci that affect susceptibility to alcoholism. We expect to begin selecting and screening candidate genes from these regions for allelic association with alcoholism related traits during the next year using the infrastructure that has been developed during the last year.

Table1. Candidate genes currently under investigation.

Symbol	Name	Chr	Size(kb)	SNPs
AC2	Adenylate cyclase type II	5	42.8	48
ADORA2a	Adenosine A2a receptor	22	9.2	12
ADCYAP1	Adenylate cyclase activating polypeptide	18	5.7	11
ADH1C	Alcohol dehydrogenase 1C subunit	4	16.3	15
CHRNA4	Nicotinic acetylcholine recptor α4 subunit	20	14.5	14
CHRNB2	Nicotinic acetylcholine receptor β2 subunit	1	8.8	10
DBH	Dopamine β hydroxylase	9	23.1	16
DRD4	Dopamine receptor D4	11	3.4	6
ENT1	Equilibrative nucleoside transporter	6	15.5	13
GNB1	G-protein β1 polypeptide	1	106.2	22
GNG2	G-protein γ2 polypeptide	14	107.8	28
NPYR2	Neuropeptide Y receptor type 2	4	8.1	10
NPYR3	Neuropeptide Y receptor type 3	2	3.9	7
KCNMA1	Potassium large conductance calcium-activated channel	10	751.5	63
OPRD1	Delta-opioid receptor	1	51.5	18
OPRK1	Kappa-opioid receptor	8	23.1	12
OPRM1	Mu-opioid receptor	6	80.1	25
PENK	Proenkephalin	8	5.7	8
POMC	Proopiomelanocortin	2	7.7	7
PRKAR1A	cAMP-dependent protein kinase type I (α-subunit)	17	39.4	18
PRKAR2B	cAMP-dependent protein kinase type II (β-subunit)	7	117.1	21
NPYR1/5	Neuropeptide Y receptor types 1 and 5	4	46.0	5

Key Research Accomplishments

- Database development so that all major processes in the laboratory, including experiment organization, sample processing and data collection, are interfaced with the database
- Bar-coding of all blood specimens, DNA samples, racks and storage boxes, and assay plates for accurate sample tracking; the database is updated automatically upon sample scanning during processing to limit data entry errors
- Automation of all liquid handling calculations and steps involved in DNA concentration determination and dilution processing. Automation of these steps was especially important in order to avoid repetitive motion injuries common with the manual pipetting steps involved in these processes
- Automation of data input from DNA processing steps into the database and label production; this limits data entry error
- Bioinformatics tools generated to facilitate experimental organization, including tools that: identify the conserved sequences between species; identify and enter into database all Celera and publicly held information on genes of interest; identify all SNPs, their location, validation status and allele frequency
- Tool to facilitate the transfer of linkage analysis conclusions into experimental follow-up with association analysis; this tool identifies and enters all human genes between linkage marker locations into the database for candidate gene selection/prioritization.
- Over 467,000 SNP genotypes collected and analyzed for 389 different polymorphisms from 23 candidate genes (>1200 subjects) by the end of November 2002
- Mendel, SimWalk2 and SOLAR statistical genetics course taken at UCLA (9/2-9/8/02) to facilitate identification of heritable phenotypic traits for application in association analysis

Reportable Outcomes

All database infrastructure and bioinformatics tools will be scalable and applicable to many other genetics/genomics applications.

Conclusions

This is a work in progress. Primary association analysis of the genotypes currently being collected and secondary exploratory analysis of phenotypic traits will be completed and reported soon.